Journal of Scientific & Industrial Research
Vol. 79, May 2020, pp. 406-412

Network Driven Discovery of Dopamine, Serotonin and Glutamate Receptors as Key Players in Schizophrenia

Sushma Rani Martha[1,2] and Manorama Patri[1]*

[1]Neurobiology Laboratory, Department of Zoology, School of Life Sciences, Ravenshaw University Cuttack, India

[2]Department of Bioinformatics, Odisha University of Agriculture and Technology, Bhubaneswar, India

Received 24 April 2019; revised 30 November 2019; accepted 4 February 2020

In the past, it was comparatively easier to mine and discover the key player genes and proteins associated with a particular disorder through literature survey. But in recent years the scientific research has attained its greatest pace resulting in the deposition of enormous molecular data over World Wide Web. Therefore, it is necessary to adopt specialized data mining methodologies and employ appropriate tools to reduce the ambiguous use of resources. In the present study, an attempt has been made to discover the most important genes and proteins involved in the Schizophrenia from all that have been reported till date by computing Average Normalized Database (AvNrD) scores of the genes. During the process, Dopamine receptor2 (DRD2), Serotonin receptor 2A (HTR2A) and Glutamate receptor 3 (GRM3) are found to be the most important proteins with significantly high AvNrD scores of 9.436725, 7.121672 and 7.709488 respectively out of 1229 genes which are likely to be associated with Schizophrenia. Analysis of the filtered gene set reveals that most of the proteins possess specific G Protein Coupled Receptor (GPCR) signature sequences. DRD and HTR proteins consist of a long Intracellular 3 (IC3) region and GLU proteins consist of an extra N-terminal Extracellular (EC) domain.

Keywords: Schizophrenia, Dopamine receptor, Serotonin receptor, Glutamate receptor, G protein coupled receptor, Network analysis

Introduction

Researches of several years have been focused in eradicating the diseases related to both physical as well as mental wellbeing which have led to tremendous important findings and hypothesis shedding light towards their diagnosis and treatment. Even though a greater degree of mystery has been solved for several disorders, disease like Schizophrenia seeks more profound and deeper research. Such problems exist due to the unpredictable phenotypic expressions of the disorder that lead to its unclear diagnosis on physiological as well as genetic basis.[1] For this reason, the treatments are aimed at symptomatic relief rather than a complete cure. Schizophrenia patients are mainly diagnosed with hallucination, loss of focus, social distraction, confusion and sometimes aggression and suicidal tendency.[2] In the present work, it is attempted to implement a novel experimental design through which all possible molecular level data could be explored on the complex brain disorder and thereafter the most important candidate genes could be recognized from the whole dataset obtained by merging the gene lists collected from 8 different databases. At the same time, we have tried to analyze the genes before and after each level of screening to evaluate their collective and comparative structural and functional aspects and interestingly arrived at some distinct findings that can be taken care of while dealing with further investigations on Schizophrenia.

Materials and Methods

Dataset preparation

All possible genes involved in Schizophrenia were extracted from 8 different most widely used databases providing molecular information on any disorder based on worldwide publications and their internal cut off parameters. GWAS[3] and GLAD4U[4] provide gene list of any particular disorder based on the publications available through PUBMED. UniProt[5] is a database containing all information on protein sequences from different organisms. PharmGKB[6], DisGeNET[7], GenAtlas[8], and GenCards[9] are databases of human genetic disorders. DrugBank[10] is the database hosting drug data along with the corresponding target protein information. All these subsets of genes were merged to

*Author for Correspondence
E-mail: mpatri@ravenshawuniversity.ac.in

form the superset containing all possible genes involved directly or indirectly with schizophrenia.

Preparation of Reduced Dataset

The first level of screening is applied to the datasets obtained from each database separately through STRING[11] network analysis. The number of 2nd order genes linked to each individual 1st order gene was listed and only those genes with **X > N where N = max(n)/3** were selected for further consideration. Here 'n' is the Number of 2nd order genes networked to each 1st order gene and 'X' is the 'n' bellow which genes were excluded from each dataset. These filtered datasets are then merged. The second level of screening is executed by considering the scores obtained by normalizing the weightage scores of the individual databases. The average of these values was computed for each gene to obtain the AvNrD scores.

Functional Classification and Domain prediction

The first level filtered and merged dataset was made to undergo functional classification using the PANTHER[12] database. This classification is based on the molecular functions and cellular localization of the proteins. GPCR regions present in the proteins were identified using PROSITE.[13] PSIPRED[14] and TMpred[15] were used to establish the secondary structural elements and transmembrane boundaries after the second level of screening. Multiple Sequence Analysis of the GPCR signature region within similar proteins was done using CLUSTAL Omega[16] through which the critical residues conserved in the case of the protein families could be identified.

Three Dimensional structural analysis

3D crystallographic structures of DRD2 (PDBID: 6CM4), HTR2A (PDBID: 6a93), GRM3 (PDBID: 4XAR) were extracted from PDB. As the structure of GRM available through PUBMED is incomplete, PHYRE 2[17] server was used to predict the 3D structure of the complete GRM3 sequence. PHYRE 2is an automated tool for 3D structure prediction of proteins based on fold recognition. Galaxy refine[18] was used for obtaining refined structures with comparatively more residues in the allowed region of Ramachandran plot. Predicted structures are validated using Procheck[19] and visualized with Discovery Studio[20] respectively.

Results and Discussion

Preparation of the dataset

A total of 1229 genes were obtained after merging the gene sets from eight different databases and removing data redundancy. Number of genes obtained from GWAS after taking P value $\leq 5 * 10^{-8}$is 399, from UniProt after restricting our search result to manually curated records only is 142, from GenAtlas after reducing our search results to records with valid Uniprot Ids is 200, from DisGeNET and GLAD4U with a cut off score >= 1 is 259 and 417 respectively, from GenCards and PharmaKb is 231 and 366 respectively and from DrugBank after downloading 80 drugs of Schizophrenia and finding their corresponding target proteins is 233 (Fig. 1).

First tier screening of the genes

By performing STRING network analysis of the gene sets of each database, we could clearly

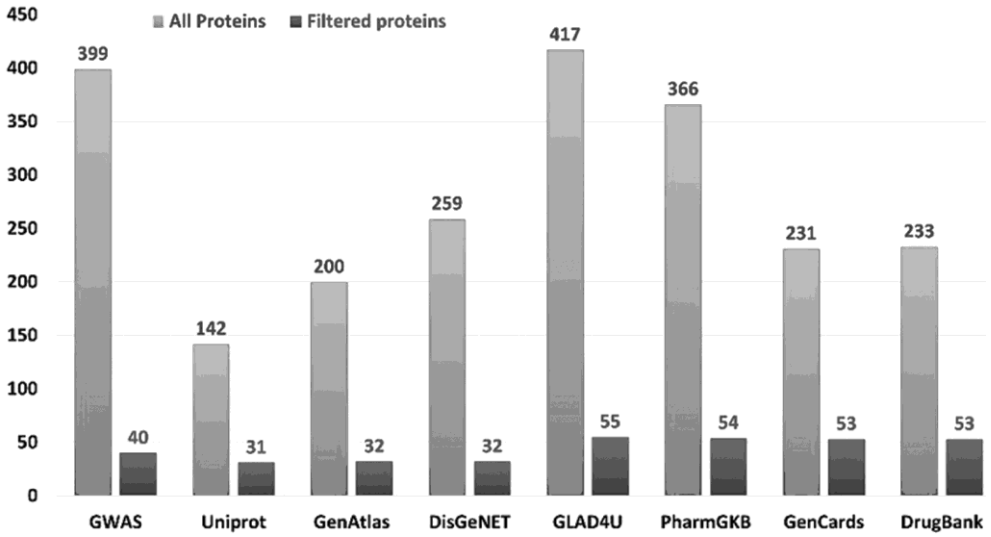

Fig. 1 — Number of genes in Complete and 1st tier filtered gene sets of different databases

distinguish between the genes highly associated with Schizophrenia from that of less associated genes remaining as outliers in the network. After screening, we got 40, 31, 32, 32, 55, 54, 53 and 53 genes from GWAS, Uniprot, GenAtlas, DisGeNET, GLAD4U, PharmGKB, GenCards, and DrugBank respectively. The reduction in the dataset is clearly depicted in the graph shown in Fig. 1. Merging all these reduced gene sets created a subset of 166 genes.

Protein classification

While trying to classify the dataset generated in the previous step into different categories, it is found that the maximum number of proteins corresponding to the genes fall into the Receptor protein category displaying the highest peak in the graph shown in Fig. 2 followed by the Transporter protein category. 22 genes code for proteins falling into the Receptor category followed by 16 genes coding for Transporter proteins. Further analysis of the receptor protein category revealed that most of the proteins in this category are GPCR proteins involved in the interaction with neurotransmitters.

Second Tier screening of genes

The scores representing the degree of relatedness of different genes/proteins with Schizophrenia are collected for each gene from GLAD4U, DisGeNET, GenCards, and DrugBank databases. A graph is plotted after computing AvNrD values within the weightage scores of the four databases. As shown in Fig. 3, DRD2 followed by HTR2A followed by GRM3 with an AvNrD score of 9.437, 7.709 and

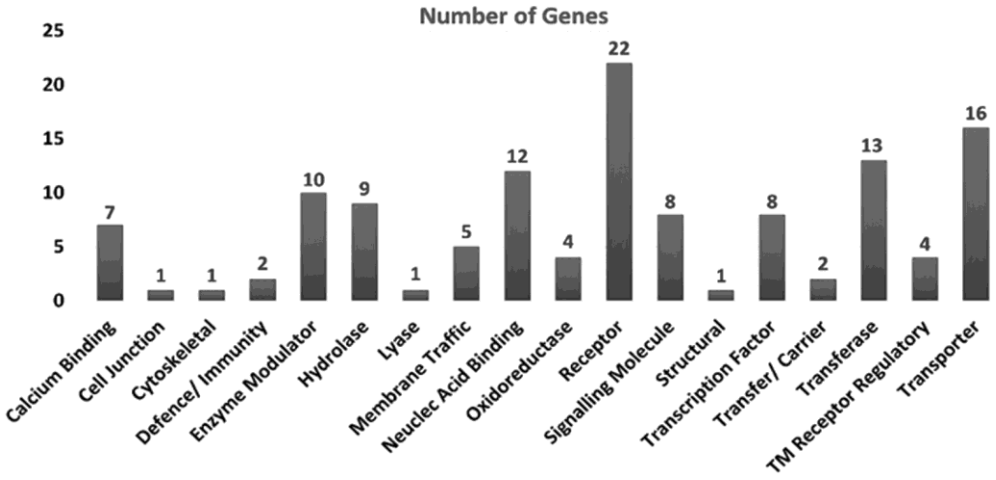

Fig. 2 — Classification of genes based on their molecular function and cellular localization of their proteins. Graph is generated using PANTHER database

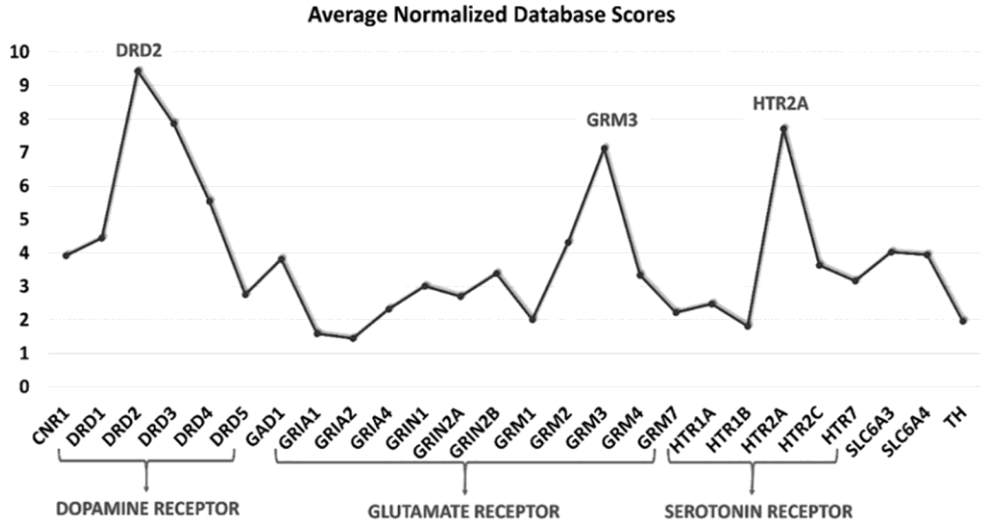

Fig. 3 — 2nd tier filtered genes/ proteins labeled in the graph: Dopamine, Serotonin & Glutamate receptors are found to be most important proteins with DRD2, HTR2A & GRM3 having significantly high score

7.122 respectively are plotted as the most important genes strongly associated with Schizophrenia due to their highest peaks in the graph. This dataset contained 26 genes that includes all five Dopamine receptor protein coding genes (DRD1, DRD2, DRD3, DRD4, DRD5), five out of 17 Serotonin receptor protein coding genes (HTR1A, HTR1B, HTR2A, HTR2C, HTR7) and eleven out of 26 Glutamate receptor protein coding genes (GRIA1, GRIA2, GRIA4, GRIN1, GRIN2A, GRIN2B, GRM1, GRM2, GRM3, GRM4, GRM7). Within these three (GRIA1, GRIA2, GRIA4) out of 4are glutamate receptors ionotropic; AMPA, three (GRIN1, GRIN2A, GRIN2B) out of 7 are glutamate receptors ionotropic; NMDA and five (GRM1, GRM2, GRM3, GRM4, GRM7) out of 8 are metabotropic glutamate receptors. Apart from these, there are other five genes named as cannabinoid receptor1 (CNR1), glutamate decarboxylase1 (GAD1), sodium depended dopamine transporter (SLC6A3), sodium depended serotonin transporter (SLC6A4) and tyrosine 3-monooxygenase (TH) also found within the group (Fig. 3).

GPCR and Transmembrane domain prediction

Analyzing the conserved domains present in all the 26 proteins revealed that 18 out of 26 proteins are having G-protein coupled receptor protein (GPCR) domain within them consisting of seven Transmembranes (TM). Table 1 summarizes all the GPCR features present and absent in the protein sequences. One disulfide bond is present in the case of dopamine and serotonin receptors which is completely absent in glutamate receptors and SLC receptors. The GPCR spanning region is longest in SLC proteins ranging between 524 aa to 526 aa whereas it is smallest in case of GRM proteins ranging between 262 aa to 274 aa. The GPCR range in HTR proteins is 289 aa to 347 aa and that of DRD proteins is 280 aa to 397 aa.

As demonstrated in Fig. 4, the conserved GPCR signature substring in DRD proteins is **ASI(f/l)NLC (A/V)IS(V/I)DR(F/Y)(V/W/T)a(V/I)**. TM1, TM2, TM3, TM4, TM5, TM6 and TM7 of DRD2 falls between 37 to 61, 71 to 98, 109 to 131, 152 to 174, 187 to 216, 374 to 398 and 408 to 429 amino acid positions. Cys-Cys disulfide bond is present between IC1 and IC2 in DRD2 (Fig. 4). Similarly the GPCR signature sequence in HTR proteins as predicted using MSA of Clustal omega is **(A/S)SI(m/l)(T/H) LC(V/A)I(S/A)(I/L)DRY(L/W/V)(g/a)I** and in GRM proteins is **FNEAK(y/f/p)I(A/G)FTM.** The SLC proteins comprise of two signature substrings that is **SFT(t/d)(e/q)LPW(i/t)(h/s)C(n/k)nswN(S/T)** and **WRFPY(l/i)cy(k/q)NGGGaF** respectively.

Three Dimensional structures of DRD2, HTR2A, and GRM3

The crystal structure recombinant DRD2 with T4 Lysozyme (DRD2R-T4L) construct is present in PDB

Table 1 — GPCR region prediction using PROSITE

Proteins	GPCR Range	Disulfide bridge	Signature Range	Signature sequence
CNR1	133–397	189–264	202–218	ASVgSLFLTAIDRYIsI
DRD1	40–331	96–186	109–125	ASIlNLCVISVDRYWaI
DRD2	51–426	107–182	120–136	ASIlNLCAISIDRYTaV
DRD3	46–383	103–181	116–132	ASIlNLCAISIDRYTaV
DRD4	51–448	108–185	121–137	ASIfNLCAISVDRFVaV
DRD5	57–359	113–217	126–142	ASIlNLCVISVDRYWaI
GRM1	592–854	Absent	781–791	FNEAKyIAFTM
GRM2	567–833	Absent	756–766	FNEAKfIGFTM
GRM3	576–842	Absent	765–775	FNEAKfIGFTM
GRM4	587–861	Absent	781–791	FNEAKpIGFTM
GRM7	590–864	Absent	784–794	FNEAKpIGFTM
HTR1A	53–400	109–187	122–138	SSIlHLCAIALDRYWaI
HTR1B	66–369	122–199	135–151	ASIlHLCVIALDRYWaI
HTR2A	91–380	148–227	161–177	ASImHLCAISLDRYVaI
HTR2C	70–368	127–207	140–156	ASImHLCAISLDRYVaI
HTR7	98–384	155–231	168–184	ASImTLCVISIDRYLgI
SLC6A3	59–585	Absent	84–98	WRFPYlcykNGGGaF
			166–186	YLfsSFTteLPWihCnnswNS
SLC6A4	78–602	Absent	103–117	WRFPYicyqNGGGaF
			186–206	YLisSFTdqLPWtsCknswNT

(6CM4: 2.87 Å resolution). The engineered sequence of this construct is manually modified at several positions to achieve a higher degree of expression, purification and crystallization. The wild type sequence is truncated from 1 to 34 position. Amino acids Isoleucine at position 22 and leucine at position 375 & 379 are mutated to Alanine [I122A; L375A; L379A]. Stretch of wild type sequence starting from Valine at 223 position up toArginineat 380 position [V223-R361] is replaced by 159 amino acid long T4 lysozyme sequence [T4L(2-161)] (Fig. 5).

The structure of recombinant HTR2A with mutated Apo cytochrome of *E coli* (HTR2AR-mbIIG) is available at PDB (6a93: 3.0 Å resolution). This structure is obtained by deleting 1 to 69 amino acids and 404 to 471 amino acids from N and C terminals respectively, by performing Serine to Lysine at 162 position and Methionine to Tryptophan [S162K; M164W] thermostabilizing mutations and by replacingthe sequence from Threonine at position 266 up to Methionine at position 312 [T266-M312] with mutated sequence of b562IIG. The mutated b56IIG is generated by mutating Methionine to Tryptophan at position 7, Arginine to Isoleucine at position 98, Histidine to Isoleucine at position 102, Arginine to Glycineat position 106 [M7W; R98I; H102I; R106G]

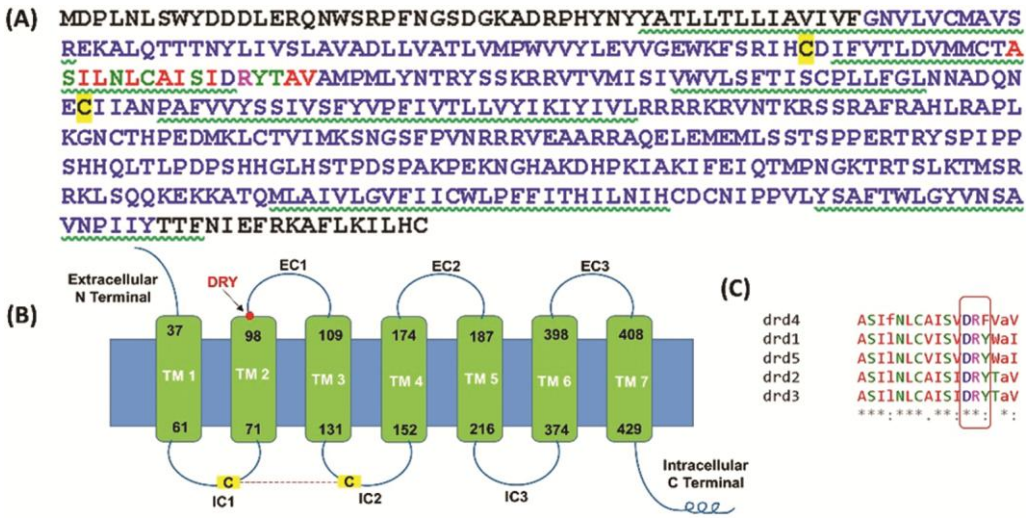

Fig. 4 — GPCR and Transmembrane domain analysis of DRD2: A. Amino acid sequence comprising the GPCR region are marked in blue; Transmembrane regions are underlined in green; GPCR signature sequence patch is shown in multiple color; Cys residues forming disulfide bond are highlighted in yellow. B. Cartoon figure indicating the transmembrane boundaries are shown along with the other conserved sites and features. C. GPCR signature sequence conserved throughout the family of the protein is shown through MSA

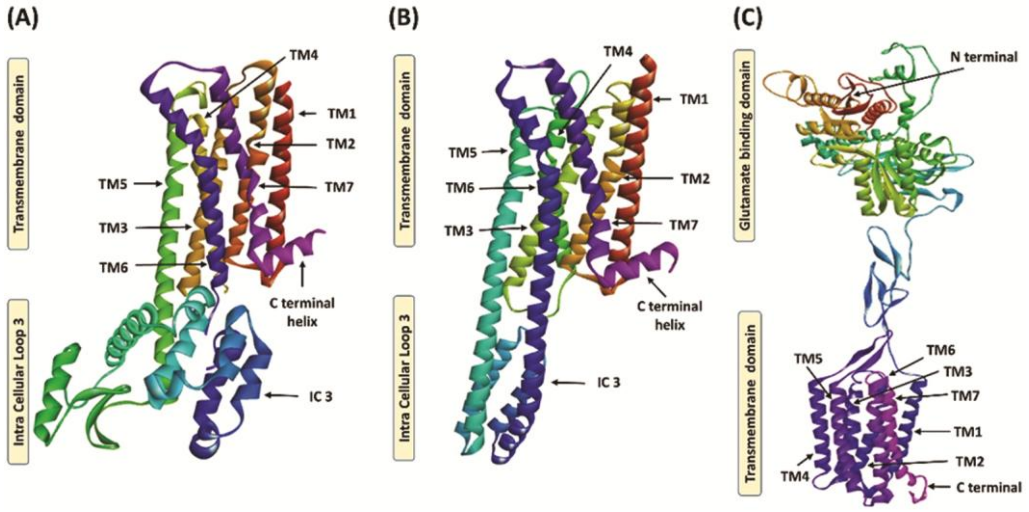

Fig. 5 — **A.** 3D structures of recombinant Dopamine receptor with T4 Lysozyme (DRD2R-T4L) **B.** 3D structures of recombinant serotonin receptor 2A with mutated Apo cytochrome of HTR2AR-mbIIG. **C.** 3D structures of GRM3 predicted using Phyre2

and by replacing loop region from Glutamine at 41 position to Phenylalanine at position 62 [Q41-F62] by [Gly-Ser-Gly-Ser-Gly] (Fig. 5).

The 3D structure of the extracellular catalytic domain of GRM3 is available at Protein Data Bank (4XAR: 2.6 Å resolution). The sequence of this PDB structure spans only between 2 to 508 on the wild type sequence. Its N terminal consists of extra Methionine, Alanine and Lysine [MAK] and C terminal consists of Histidine tag (Gly-His-His-His-His-His-His). The complete structure of GRM3 predicted through Phyre2 and refined through Galaxy Refine is placed in Fig. 5c, where the two domains (one glutamate binding domain or extracellular catalytic domain and another Transmembrane GPCR domain) are distinctly labeled. This sequence spans between 30 to 831 position on wild type sequence out of which 1 to 510 forms glutamate binding domain and 565 to 831 forms the GPCR domain. The complete structure of GRM3 is having 92.4% of amino acid residues in the allowed region, 7.2% in the additionally allowed region, 0.4% (Ala487, His498, Val798) in the generously allowed region and 0.0% in the disallowed region of Ramachandran plot.

Conclusion

The complete work is dedicated to reveal the major key player genes and thence their corresponding proteins involved in schizophrenia and make an analysis of these gene sets to reveal and recognize if any common and/or specific structural and functional features are shared by them. After applying all prediction methods, we could clearly observe that the majority of the genes code for Transmembrane GPCR proteins involved in the regulation and movement of neurotransmitters. The majority of the proteins have a short GPCR signature sequence pattern conserved within the protein families. Along with 7 TM helices that are embedded within the cell membrane in case of most of the proteins, the dopamine receptor proteins and serotonin receptor proteins have a long variable IC3 region whereas glutamate receptor proteins have a large extracellular N-terminal region. Both DRD and HTR proteins have a DRY(Asp-Arg-Tyr) conserved sites and a disulfide Cys-Cys double bond which is absent in GRM proteins. DRD2, HTR2A and GRM3 are found to have significantly higher AvNrD score within the 26 genes that have been identified after two levels of screening. The 3-D structures of the GPCR regions of the proteins were found and refined to achieve maximum accuracy. Effect of mutations on the structural and functional stabilities of these proteins and their corresponding contribution towards the cause of the disorder along with the study of the Schizophrenia drugs interacting with the proteins are under our concern for our upcoming research works.

Acknowledgment

The present study was carried by the authors with the support of the Life Sciences Department, Ravenshaw University, Cuttack and Bioinformatics Department, Orissa University of Agriculture and Technology, Bhubaneswar.

References

1 Moustafa A A, Phillips J, Keri S, Misiak B & Frydecka D, On the Complexity of Brain Disorders: A Symptom-Based Approach, *Front Comput Neurosci*, (2016) doi:10.3389/ fncom.2016.00016.

2 Craddock N, O' Donovan M C & Owen M J, The genetics of schizophrenia and bipolar disorder: dissecting psychosis, *J Med Genet*, **42** (2005) 193–204.

3 *MacArthur J, Bowler E, Cerezo M,* Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, Pendlington Z M, Welter D, Burdett T, Hindorff L, Flicek P, Cunningham F & Parkinson H,The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog), *Nucleic Acids Res*, **45** (2017) D896–D901.

4 Jourquin J, Duncan D, Shi Z & Zhang B, GLAD4U: deriving and prioritizing gene lists from PubMed literature, *BMC Genomics*, **201213** (2012) (Suppl 8) S20.

5 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge A J, Poux S, Bougueleret L & Xenarios I, UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View, *Methods Mol Biol*, **1374** (2016) 23–54.

6 Hicks J K, Bishop J R, Sangkuhl K, Muller D J, Ji Y, Leckband S G, Leeder J S, Graham R L, Chiulli D L, LLerena A, Skaar T C, Scott S A, Stingl J C, Klein T E, Caudle K E & Gaedigk A, Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors, *Clin Pharmacol Ther*, **98(2)** (2015)127–134.

7 Bauer-Mehren A, Rautschka M, Sanz F & Furlong L I, DisGeNET: A Cytoscape plugin to visualize, integrate, search and analyze gene–disease networks, *Bioinformatics*, **26(22)** (2010) 2924–2926.

8 Frezal J, Genatlas database, genes and development defects, *C R Acad Sci III*, **321(10)** (1998) 805–817.

9 Rebhan M, Chalifa-Caspi V, Prilusky J & Lancet D, GeneCards: integrating information about genes, proteins and diseases, *Trends Genet*, **13(4)** (1997) 163.

10 Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo A C& Wishart D S, DrugBank 3.0: a comprehensive resource for 'omics' research on drugs, *Nucleic Acids Res*, **39** (2011) D1035–D1041.

11 Szklarczyk D, Franceschini A, Wyder, S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou K P, Kuhn M, Bork P, Jensen L J & von Mering C, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res*, **43** (2015) D447–D452.

12 Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas P D, PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium, *Nucleic Acids Res*, **38** (2010) D204-D210.

13 Sigrist C J A, de Castro E, Cerutti L, Cuche B A, Hulo N, Bridge A, Bougueleret L& Xenarios I, New and continuing developments at PROSITE, *Nucleic Acids Res*, **D1** (2012) D344–D347.

14 Buchan D W A, Minneci F, Nugent T C, Bryson K & Jones D T, Scalable web services for the PSIPRED Protein Analysis Workbench, *Nucleic Acids Research*, **41(W1)** (2013) W340–W348.

15 Hofmann K & Stoffel W, TMbase - A database of membrane spanning proteins segments, *Biol Chem Hoppe-Seyler* **374** (1993)166.

16 Corpet, F, Multiple sequence alignment with hierarchical clustering, *Nucl Acids Res*, **16** (1988) (22) 10881–10890

17 Kelley L A, Mezulis S, Yates C M, Wass M N & Sternberg M J, The Phyre2 web portal for protein modeling, prediction and analysis, *Nature Protocols*, **10** (2015) 845–858.

18 Heo L, Park H & Seok C, Galaxy Refine: protein structure refinement driven by side-chain repacking, *Nucleic Acids Res*, **41** (2013) W384–W388.

19 Laskowski R A, Rullmannn J A, MacArthur M W, Kaptein R & Thornton J M, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR*, **8** (1996) 477–486.

20 *Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.*